

# Set-based Multiobjective Fitness Landscapes: A Preliminary Study

Sébastien Verel  
Univ. Nice Sophia-Antipolis  
INRIA Lille  
verel@i3s.unice.fr

Arnaud Liefooghe  
Université Lille 1  
LIFL – CNRS – INRIA Lille  
arnaud.liefooghe@lifl.fr

Clarisse Dhaenens  
Université Lille 1  
LIFL – CNRS – INRIA Lille  
clarisse.dhaenens@lifl.fr

## ABSTRACT

Fitness landscape analysis aims to understand the geometry of a given optimization problem in order to design more efficient search algorithms. However, there is a very little knowledge on the landscape of multiobjective problems. In this work, following a recent proposal by Zitzler *et al.* (2010), we consider multiobjective optimization as a set problem. Then, we give a general definition of set-based multiobjective fitness landscapes. An experimental set-based fitness landscape analysis is conducted on the multiobjective  $NK$ -landscapes with objective correlation. The aim is to adapt and to enhance the comprehensive design of set-based multiobjective search approaches, motivated by an *a priori* analysis of the corresponding set problem properties.

## Categories and Subject Descriptors

F.2.m [Analysis of Algorithms and Problem Complexity]: Miscellaneous

## General Terms

Algorithms

## Keywords

Fitness landscapes, Multiobjective optimization, Set-based multiobjective search

## 1. INTRODUCTION

There exists a large amount of literature about multiobjective optimization in general, and about the identification or the approximation of the Pareto optimal set in particular. In the latter case, evolutionary multiobjective optimization (EMO) techniques have received a growing interest since the late 1980s. The overall goal is generally to identify a set of good-quality solutions (ideally the whole or a ‘representative’ subset of the Pareto optimal set). As a consequence, recent advances in the field explicitly formulate the goal of multiobjective optimization as a set problem:

the search space is made of sets of solutions (and not single solutions) [17]. However, to date, the impact of the main problem-related properties on the behavior and the performance of set-based multiobjective search approaches is still far from being well-understood.

Up to now, the definition of multiobjective fitness landscapes (moFiL) has been mainly restricted upon two different levels: the properties of the Pareto optimal set on the one hand, and of the search space properties at the solution-level on the other hand. With respect to the Pareto optimal set, problem-related properties are known to largely affect the structure of Pareto optimal solutions [10], and then the behavior of search algorithms [11]. With respect to the solution-level, Knowles and Corne [8] lead a landscape analysis on the multiobjective quadratic assignment problem with a rough objective correlation. The transposition of standard tools from fitness landscape analysis to multiobjective optimization are discussed by Garrett [4], together with a study on fitness-distance correlation. In another study, a moFiL is regarded as a neutral landscape, and divided into different fronts with the same dominance rank [5]. In previous works on multiobjective  $NK$ -landscapes [1], enumerable moFiL are studied according to the number of fronts, the number of solutions on each front, the probability to pass from one front to another, and the hypervolume-value of the Pareto optimal set.

These types of moFiL lead to rather poor tools to describe the dynamics of population-based multiobjective search algorithms. We here propose to define a third type of moFiL, dealing with the *search space properties at the set-level*. The contributions of this work are summarized below.

- (i) A definition of *set-based multiobjective fitness landscapes* is given, based on a search space made of solution-sets, a neighborhood relation between solution-sets, and an indicator-based fitness function.
- (ii) An experimental analysis is conducted in order to study standard tools from single-objective fitness landscapes (ruggedness and multimodality) in the context of set-based multiobjective search. We study the influence of the main problem-related properties and of the solution-set size on multiobjective  $NK$ -landscapes.

The reminder of the paper is organized as follows. Section 2 deals with fitness landscapes, multiobjective optimization, and set-based multiobjective search. In section 3, we give a definition of set-based multiobjective fitness landscapes, illustrated with multiple examples. Experimental results are given in Section 4; and the last section concludes the paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO’11, July 12–16, 2011, Dublin, Ireland.

Copyright 2011 ACM 978-1-4503-0557-0/11/07 ...\$10.00.

## 2. PRELIMINARIES

### 2.1 Fitness Landscapes

In single-objective optimization, the notion of fitness landscape (FiL) has been introduced to study the topology of a problem [6]. A FiL can be defined by the triplet  $(S, \mathcal{N}, f)$  such that:  $S$  is a set of admissible solutions (*i.e.* the search space);  $\mathcal{N} : S \rightarrow 2^S$  is a *neighborhood relation*, *i.e.* a function that assigns a set of solutions  $\mathcal{N}(s) \subset S$  to any solution  $s \in S$ , the set  $\mathcal{N}(s)$  is called the *neighborhood* of  $s$ , and a solution  $s' \in \mathcal{N}(s)$  is called a *neighbor* of  $s$ ;  $f : S \rightarrow \mathbb{R}$  is a fitness function that can be pictured as the ‘height’ of the corresponding solutions, here assumed to be maximized. A *local optimum* is a solution  $s^* \in S$  such that  $\forall s \in \mathcal{N}(s^*)$ ,  $f(s) \leq f(s^*)$ . The ability of search algorithms is related to the number of local optima, and to their distribution over the landscape [9]. *Global optima* are defined as the absolute maxima in the whole search space  $S$ . Other landscape features, such as basins, barriers, or neutrality can also be defined [12]. For the sake of self-containedness, several notions that will be used later in the paper are defined below.

A *walk* on the landscape from  $s$  to  $s' \in S$  is a sequence  $(s_0, s_1, \dots, s_m)$  of solutions from the search space such that  $s_0 = s$ ,  $s_m = s'$  and  $s_i \in \mathcal{N}(s_{i-1}) \forall i \in \{1, \dots, m\}$ . For instance, the walk is said to be random if solutions are chosen with a uniform probability from the neighborhood. It can also be obtained through the repeated application of a ‘move’ operator defined on the landscape, such as a random mutation or a deterministic hill-climbing.

Given a walk  $(s_t, s_{t+1}, \dots)$ , the *autocorrelation function* [14] ( $\rho$ ) of a fitness function  $f$  is the autocorrelation function of time series  $(f(s_t), f(s_{t+1}), \dots)$ :

$$\rho(k) = \frac{E[f(s_t)f(s_{t+k})] - E[f(s_t)]E[f(s_{t+k})]}{\text{var}(f(s_t))}$$

where  $E[f(s_t)]$  and  $\text{var}(f(s_t))$  are the expected value and the variance of  $f(s_t)$ , respectively. Estimates  $r(k)$  of autocorrelation coefficients  $\rho(k)$  can be calculated with a time series  $(s_1, s_2, \dots, s_L)$  of length  $L$ :

$$r(k) = \frac{\sum_{j=1}^{L-k} (f(s_j) - \bar{f})(f(s_{j+k}) - \bar{f})}{\sum_{j=1}^L (f(s_j) - \bar{f})^2}$$

where  $\bar{f} = \frac{1}{L} \sum_{j=1}^L f(s_j)$ , and  $L \gg 0$ . A random walk is representative of the landscape when it is statistically isotropic. In such a case, whatever the starting point and the neighbors selected during the walks, estimates of  $r(n)$  are nearly the same. The estimation error diminishes with the length of the walk. The *autocorrelation length*  $\tau$  measures how the autocorrelation function decreases. This summarizes the ruggedness of the landscape: the larger the correlation length, the smoother the landscape. Weinberger’s definition  $\tau = -\frac{1}{\ln(\rho(1))}$  makes the assumption that the autocorrelation function decreases exponentially [14].

The length of *adaptive walks*, performed with a hill-climber, is an estimator of the diameter of the local optima basins of attraction. The larger the length, the larger the basin size. This allows to estimate the number of local optima when the whole search space cannot be enumerated exhaustively.

### 2.2 Multiobjective Optimization

A multiobjective optimization problem can be defined by a set of  $M \geq 2$  objective functions  $F = (f_1, f_2, \dots, f_M)$ , and

a set  $\mathcal{X}$  of feasible solutions in the *decision space*. In the combinatorial case,  $\mathcal{X}$  is a discrete set. Let  $\mathcal{Z} = F(\mathcal{X}) \subseteq \mathbb{R}^M$  be the set of feasible outcome vectors in the *objective space*. In a maximization context, a solution  $x' \in \mathcal{X}$  is dominated by a solution  $x \in \mathcal{X}$ , denoted by  $x' \prec x$ , iff  $\forall i \in \{1, 2, \dots, M\}$ ,  $f_i(x') \leq f_i(x)$  and  $\exists j \in \{1, 2, \dots, M\}$  such that  $f_j(x') < f_j(x)$ . A solution  $x^* \in \mathcal{X}$  is said to be *Pareto optimal* (or *efficient*, *non-dominated*), if there does not exist any other solution  $x \in \mathcal{X}$  such that  $x^* \prec x$ . The set of all Pareto optimal solutions is called the *Pareto optimal set* (or the *efficient set*). Its mapping in the objective space is called the *Pareto front*. A common approach is to identify a minimal complete Pareto optimal set, for which each point of the Pareto front corresponds to a single Pareto optimal solution.

However, generating the entire Pareto optimal set is often infeasible for two main reasons: (i) the number of Pareto optimal solutions is typically exponential in the size of the problem instance, and (ii) deciding if a feasible solution belongs to the Pareto optimal set is often NP-complete. Therefore, the overall goal is often to identify a good *Pareto set approximation*. To this end, evolutionary algorithms have received a growing interest since the late eighties.

### 2.3 Set-based Multiobjective Search

Recently, approximating the Pareto optimal set has been explicitly stated as a *set problem* [17]. In that sense, most existing EMO algorithms can be seen as hill-climbers performing on sets. Let us define the search space  $\Sigma \subset 2^{\mathcal{X}}$  by a set of feasible sets of solutions (and not single solutions). An element  $\sigma \in \Sigma$  is denoted as a *solution-set*. Usually, a maximum cardinality is imposed:  $|\sigma| \leq \mu$  for all  $\sigma \in \Sigma$ . Different interpretations of what is a good Pareto set approximation are possible, and the definition of approximation quality strongly depends on the decision-maker preferences. A *set preference relation* is then usually induced over  $\Sigma$ , like the Pareto dominance relation extended to solution-sets.

We here assume that the set preference relation is explicitly given in terms of a quality indicator  $\mathbf{I} : \Sigma \rightarrow \mathbb{R}$ . One of them is the *hypervolume indicator*  $\mathbf{I}_H$  [18], that is to be maximized. It gives the portion of the objective space enclosed by a solution-set  $\sigma \in \Sigma$  and a reference point  $z^* \in \mathcal{Z}$ . The hypervolume indicator is one of the most commonly used indicator, due to several interesting properties [15]. In particular, this is the only indicator that is dominance preserving, *i.e.*  $\forall \sigma, \sigma' \in \Sigma$  such that  $\sigma'$  is dominated by  $\sigma$ :  $\mathbf{I}_H(\sigma) \geq \mathbf{I}_H(\sigma')$ . Many recent search algorithms are based on the hypervolume indicator, but most of them operates at the solution-level [3, 16], with the exception of [2]. The goal of a hypervolume-based search is then to find a solution-set  $\sigma \in \Sigma$  that maximizes the indicator value:

$$\arg \max_{\sigma \in \Sigma} \mathbf{I}_H(\sigma) \quad (1)$$

Let us note that a minimal solution-set maximizing  $\mathbf{I}_H$  is a subset of the Pareto optimal set. Therefore,  $\mathbf{I}_H$  can be seen as a function that assigns, to each solution-set, a scalar value reflecting its quality according to the goal formulated in (1), *i.e.* a *fitness function* defined over sets.

## 3. SET-BASED FITNESS LANDSCAPES

### 3.1 Definition

Like in single-objective optimization, a multiobjective fitness landscape (moFiL) requires a proper definition of (i) a

search space, (ii) a neighborhood operator, and (iii) a fitness function. From a multiobjective perspective, several remarks and criticisms can be stated from previous attempts made in the past in defining a moFiL. First, the output of a multiobjective search algorithm is a solution-set, and not a single solution like in the single-objective case. Moreover, multiobjective search approaches in general manipulate either a population of solutions, or an archive of mutually non-dominated solutions. Both can be viewed as solution-sets. As a consequence, following the work of Zitzler et al. [17], identifying multiple tradeoff solutions by means of a Pareto set approximation can explicitly be stated as a set problem (see Section 2.3). The search space of a multiobjective optimization problem is here assumed to be constituted of a set of feasible solution-sets. Second, considering a partial order only to analyze a FiL does not allow to measure interesting fitness landscape features dealing with the ruggedness and the evolvability (among others). This is the reason why the Pareto dominance relation (or a slight modification of it) is generally not satisfying enough to define a moFiL. Quality indicators as defined in [18] allow to overcome such a limitation by introducing a complete order between solution-sets, and by quantifying their respective quality with respect to the indicator being used. Last, in their proposal on set-based multiobjective search, Zitzler et al. [17] do not define any set-based neighborhood operator, then restricting the application of their approach to some ‘*random set mutation*’, or ‘*heuristic set mutation*’. However, defining a neighborhood structure on solution-sets allows to distinguish between the properties of the search space, and the heuristics used to explore solution-set’s neighborhood. This is also through this definition that are located the main differences in the dynamics of set-based multiobjective search algorithms.

In this work, we propose the definition of a moFiL in terms of set-based multiobjective search by means of an indicator-based fitness function.

A *set-based multiobjective fitness landscape* is defined as a triplet  $(\Sigma, \mathcal{N}, \mathcal{I})$  such that:

- $\Sigma \subset 2^{\mathcal{X}}$  is a set of feasible solution-sets (where  $\mathcal{X}$  is the set of feasible solutions);
- $\mathcal{N} : \Sigma \rightarrow 2^{\Sigma}$  is a neighborhood relation between solution-sets;
- $\mathcal{I} : \Sigma \rightarrow \mathbb{R}$  is a unary quality indicator, *i.e.* a fitness function measuring the quality of solution-sets.

$\Sigma$ ,  $\mathcal{N}$ , and  $\mathcal{I}$  still need to be defined for the problem at hand. But this is also the case in single-objective optimization, except that they are here defined at the set-level.

Algorithm 1 gives a general class of algorithms that set-based moFiL are able to compare. For sure, most existing multiobjective search algorithms can be formulated as instances of this general methodology.

### 3.2 Illustrative Examples of Set-based moFiL

Different set-level search spaces can be considered according to the problem and the algorithm under study. Several examples are given below.

- The search space of population-based approaches can

---

#### Algorithm 1 Set-based Neighborhood Search Algorithm

---

```

start with a solution-set  $\sigma \in \Sigma$ 
evaluate  $\sigma$  with respect to  $\mathcal{I}$ 
repeat
  select  $\sigma' \in \mathcal{N}(\sigma)$ 
  evaluate  $\sigma'$  with respect to  $\mathcal{I}$ 
  if  $\text{accept}(\sigma, \sigma')$  then
     $\sigma \leftarrow \sigma'$ 
  end if
until  $\text{continue}(\sigma)$ 
return non-dominated solutions of  $\sigma$ 

```

---

be defined as  $\Sigma = \{\sigma \in 2^{\mathcal{X}} : |\sigma| = \mu\}$ , where  $\mu$  is the population size.

- The search space of approaches using a bounded archive can be defined as  $\Sigma = \{\sigma \in 2^{\mathcal{X}} : |\sigma| \leq \mu\}$ , where  $\mu$  is the maximal size of the archive.
- The search space of a number of existing dominance-based approaches, where solution-sets of mutually non-dominated solutions only are considered, can be defined as  $\Sigma = \{\sigma \in 2^{\mathcal{X}} : \forall s, s' \in \sigma, s \not\prec s'\}$ .
- A search space with the two previous restrictive conditions can also be considered, *i.e.*  $\Sigma = \{\sigma \in 2^{\mathcal{X}} : |\sigma| \leq \mu \text{ and } \forall s, s' \in \sigma, s \not\prec s'\}$ .
- A search space without any restriction is  $\Sigma = 2^{\mathcal{X}}$ .

Next, the neighborhood structure has to reflect the way the search space is explored by a class of search algorithms. In the general case, the definition of neighborhood is based either on a distance, or more often on the variation operator(s) handled by the algorithm under study. Roughly speaking, at the set-level, the neighbors of a solution-set can for instance be obtained by (i) replacing a solution from the set, (ii) inserting a solution to the set, or (iii) deleting a solution from the set. In order to give more precise examples of set-level neighborhood operators, let us consider an arbitrary non-empty solution-set  $\sigma \in \Sigma$ , an arbitrary non-empty neighboring solution-set  $\sigma' = \mathcal{N}(\sigma)$ , and an arbitrary neighboring solution  $s' \in \mathcal{N}(s)$  with  $s \in \sigma$ . Possible set-level neighborhood operators are discussed below.

- When replacing a solution from the set, a neighboring solution-set can be defined as  $\sigma' = \sigma \cup \{s'\} \setminus \{s''\}$  such that  $s'' \in \sigma$ . The size of this replacement set-level neighborhood is at most  $|\sigma| \cdot \sum_{s \in \sigma} |\mathcal{N}(s)|$ .

In such a case, a possible neighborhood exploration strategy is to find the tuple  $(s', s'')$ , with  $s'' \in \sigma$ , such that  $\mathcal{I}(\sigma \cup \{s'\} \setminus \{s''\})$  is maximal. However, most existing EMO methodologies generally separate the fact of inserting a solution to the set, and deleting a solution from the set into two different phases.

- When inserting a new solution to the set, a neighboring solution-set can be defined as  $\sigma' = \sigma \cup \{s'\}$ . The size of this insertion neighborhood is at most  $\sum_{s \in \sigma} |\mathcal{N}(s)|$ .
- When deleting a solution from the set, a neighboring solution-set can be defined as  $\sigma' = \sigma \setminus \{s\}$  where  $s \in \sigma$ . The size of this deletion neighborhood is  $|\sigma|$ .

The set-level neighborhood operators can be applied multiple times in order to define large-size neighborhood operators, where several solutions can differ in a neighboring solution-set. A neighboring solution-set must always correspond to an element of the given search space. As a consequence, when the solution-sets are somehow bounded in size, the neighborhood must be restricted using a (partial) dominance relation, or a limited-size set. Of course, the definition of a set-level neighborhood relations are not limited to the use of a solution-level neighboring operator  $\mathcal{N}$ . For instance, a set-level neighborhood relation can consider a random solution, or a solution produced by applying a recombination operator to pairs of solutions in the solution-set, and so on. Anyway, all those set-level neighborhood operators are just few examples, and like in single-objective optimization, one has to define the neighborhood relation according to the (set) problem and the algorithm under study.

At last, the fitness function defined for set-based moFiL is given in terms of a quality indicator. Several studies are devoted to theoretical properties of multiobjective quality indicators. For more details, the reader is referred to [18].

### 3.3 Discussion

In the general case, two typical uses of FiL analysis can be conducted. First, such a study can allow to compare the difficulties, in terms of FiL features, associated with different search problems. Given a search algorithm and two different optimization problems, the corresponding FiL are defined (*i.e.* the search space, the neighborhood relation, and the fitness function). Then, the difficulties can be compared between both FiL according to measures dealing, for instance, with the number of local optima, their distribution, the ruggedness, the evolvability, and so on. Second, another possibility of FiL analysis is the off-line tuning or design of search approaches. Once again, given a search problem and different possible component design or parameter setting, the corresponding landscapes are defined. Then, according to the FiL measures, the most promising search algorithm components can be chosen *a priori*. In the context of set-based multiobjective search, a comparison of two set-based moFiL can be compared with each other in terms of FiL measures. They can be defined, for instance, by two different neighborhood operators, two different fitness functions or two different search space definitions. In the following, we conduct an empirical study on the comparison of difficulty of multiobjective optimization problems.

## 4. EXPERIMENTAL ANALYSIS

### 4.1 $\rho MNK$ -Landscapes

In the single-objective case, the family of  $NK$ -landscapes constitutes an interesting model to study the influence of non-linearity on the number of local optima. In this section, we present the  $\rho MNK$ -landscapes proposed in [13]. Four parameters are required to define a  $\rho MNK$ -landscape: the problem size  $N$ , the number of epistatic links  $K$ , the number of objectives  $M$ , and the objective correlation coefficient  $\rho$ .

The family of  $NK$ -landscapes is a problem-independent model used for constructing multimodal landscapes [7]. Parameter  $N$  refers to the number of bits in the decision space (*i.e.* the string length) and  $K$  to the number of bits that influence a particular bit from the string (the epistatic interactions). By increasing the value of  $K$  from 0 to  $(N - 1)$ ,  $NK$ -

landscapes can be gradually tuned from smooth to rugged. The fitness function (to be maximized) of a  $NK$ -landscape  $f_{NK} : \{0, 1\}^N \rightarrow [0, 1]$  is then defined on binary strings of size  $N$ . An ‘atom’ with a fixed epistasis level is represented by a fitness component  $f_i : \{0, 1\}^{K+1} \rightarrow [0, 1]$  associated with each bit  $i \in N$ . Its value depends on the allele at bit  $i$  and also on the alleles at  $K$  other bit positions ( $K$  must fall between 0 and  $N - 1$ ). In other words, the parameter  $K$  tunes the degree of non-linearity (epistasis). The fitness  $f_{NK}(x)$  of a solution  $x \in \{0, 1\}^N$  corresponds to the mean value of its  $N$  fitness components  $f_i$ :  $f_{NK}(x) = \frac{1}{N} \sum_{i=1}^N f_i(x_i, x_{i_1}, \dots, x_{i_K})$ , where  $\{i_1, \dots, i_K\} \subset \{1, \dots, i-1, i+1, \dots, N\}$ . In this work, we set the  $K$  bits randomly. Each fitness component  $f_i$  is specified by extension, *i.e.* a number  $y_{x_i, x_{i_1}, \dots, x_{i_K}}^i$  from  $[0, 1]$  is associated with each element  $(x_i, x_{i_1}, \dots, x_{i_K})$  from  $\{0, 1\}^{K+1}$ . Those numbers are uniformly distributed in the range  $[0, 1]$ . As a consequence, it is very unlikely that the same fitness value is assigned to two different solutions.

A multiobjective variant of  $NK$ -landscapes (namely  $MNK$ -landscapes) has been defined with a set of  $M$  independent fitness functions [1]. The same epistasis degree  $K_m = K$  is used for all the objectives. Each fitness component  $f_{m,i}$  is specified by extension with the numbers  $y_{x_i, x_{i_1}, \dots, x_{i_K}}^{m,i}$ . In the original  $MNK$ -landscapes, these numbers are defined randomly and independently. An approach for designing  $MNK$ -landscapes with correlated objective functions has been recently proposed in [13]. First, let us define the  $CMNK$ -landscapes, where the epistasis structure is identical for all the objective functions:  $\forall m \in \{1, \dots, M\}$ ,  $K_m = K$  and  $\forall m \in \{1, \dots, M\}$ ,  $\forall j \in \{1, \dots, K\}$ ,  $i_{m,j} = i_j$ . However, the fitness components are not defined independently. The numbers  $(y_{x_i, x_{i_1}, \dots, x_{i_K}}^{1,i}, \dots, y_{x_i, x_{i_1}, \dots, x_{i_K}}^{M,i})$  follow a multivariate uniform law of dimension  $M$ , defined by a correlation matrix  $C$ . Thus, the  $y$ ’s follow a multidimensional law with uniform marginals and the correlations between  $y$ ’s are defined by the matrix  $C$ . The construction of  $CMNK$ -landscapes defines correlation between the  $y$ ’s but not directly between the objectives. In [13], it is proven by algebra that the correlation between objectives is tuned by the matrix  $C$ :  $E(\text{cor}(f_n, f_p)) = c_{np}$ . In the  $\rho MNK$ -landscapes, the correlation matrix  $C_\rho = (c_{np})$  is assumed to have the same correlation between all pairs of objectives:  $c_{nn} = 1$  for all  $n$ , and  $c_{np} = \rho$  for all  $n \neq p$ . Of course, for obvious reasons, it is not possible to have the matrix  $C_\rho$  for all  $\rho$  values in  $[-1, 1]$ :  $\rho$  must be greater than  $\frac{-1}{M-1}$ , see [13]. In  $\rho MNK$ -landscape, the parameter  $\rho$  allows to tune very precisely the correlation between all pairs of objectives.

In the following, we conduct an empirical study of the influence of the problem dimension, the non-linearity (epistasis), the number of objective functions and the objective correlation on some properties of set-based moFiL.

### 4.2 Experimental Design

In order to minimize the influence of the random creation of landscapes, we considered 30 different and independent instances for each parameter combinations:  $\rho$ ,  $M$ , and  $K$ . The measures reported are the average over these 30 instances. The parameters under investigation in this study are given in Table 1. We analyze the multiobjective  $\rho MNK$ -landscapes according to set-based search algorithms that manipulate a fixed-size solution-set. The goal is to show the link be-

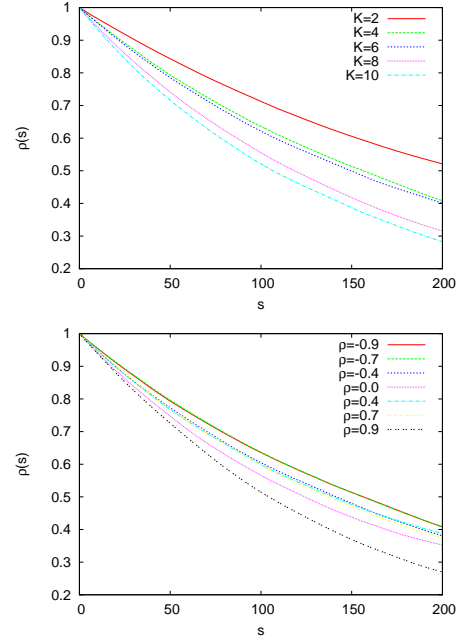
**Table 1: Parameters used in the paper.**

Parameter	Values
$N$	$\{64\}$
$M$	$\{2, 3, 5\}$
$K$	$\{2, 4, 6, 8, 10\}$
$\rho$	$\{-0.9, -0.7, -0.4, -0.2, 0.0, 0.2, 0.4, 0.7, 0.9\}$ such that $\rho \geq -1/(M-1)$

tween the geometry of the set-based moFiL and the features that make a search algorithm efficient for the corresponding problem. Previous results indicate that the problem is getting more complex when the non-linearity and the degree of conflict between the objectives are high [1, 13]. Feasible solutions are bit strings of size  $N$ :  $\mathcal{X} = \{0, 1\}^N$ , and the set-level search space is the set of solution-sets of size  $\mu$ . The set-level neighborhood relation consists of the replacement neighborhood as defined in Section 3. It does not change the solution-set size and uses a bit-flip solution-level neighborhood operator. In this work, we do not consider the possible insertion or deletion of solutions from the solution-set. Hence, two solution-sets are neighbors if they have the same size, and if they differ by one solution only. It is also required that the corresponding solutions are neighbors according to the one bit-flip neighborhood operator:  $\sigma' \in \mathcal{N}(\sigma)$  iff  $\exists s \in \sigma, \exists s' \in \mathcal{X}$  such that  $d_{\text{Hamming}}(s', s) = 1$  and  $\sigma' = \sigma \setminus \{s\} \cup \{s'\}$ . The maximal size of this neighborhood relation is then  $(|\sigma| \cdot N)$ . The set-level fitness function is based on the hypervolume indicator [18]. Given that the objective functions of  $\rho MNK$ -landscapes, defined in [0, 1], are to be maximized, the reference point required by the hypervolume calculation is set to  $0^M$ .

### 4.3 Ruggedness

The ruggedness of a multiobjective problem is here measured in terms of the autocorrelation of the hypervolume along a random walk. The starting solution-set of the walk is initialized with  $\mu = 100$  random solutions. At each step of the random walk, a random neighboring solution-set replaces the current one. The length of the random walk is set to  $5 \cdot 10^3$ . Figure 1 shows the autocorrelation functions for an objective space dimension  $M = 2$  with respect to the non-linearity degree  $K$ , and to the objective correlation  $\rho$ . The functions all decrease slowly with the step lag. The hypervolume correlation between random neighboring solution-sets is high. Figure 2 shows the autocorrelation length according to parameter  $K, M$  and  $\rho$ . The correlation values are very high. As a comparison, the autocorrelation length of single-objective  $NK$ -landscapes is  $-1/\log(1 - \frac{K+1}{N})$ , which gives the length 20.8 for  $N = 64$  and  $K = 2$  [12]. The correlation between neighboring solutions with respect to each objective function impacts the correlation between neighboring solution-sets in terms of hypervolume. But this correlation also depends on the solution-set size  $\mu$ . Let us suppose that the fitness values between neighboring solutions change with a factor  $\alpha$ . Then, the change of the hypervolume values between the corresponding neighboring solution-sets is lower than  $\alpha$ . Notice that the magnitude of the autocorrelation length relative to the hypervolume is approximately  $\mu$  times the one related to the solution-level fitness values. Nevertheless, as the well-known result from single-objective  $NK$ -landscapes, the autocorrelation length of the hyper-



**Figure 1: Autocorrelation functions according to parameter  $K$  (top  $\rho = -0.2$ ), and to parameter  $\rho$  (bottom  $K = 2$ ). The number of objectives is  $M = 2$ .**

volume decreases with the non-linearity degree of  $\rho MNK$ -landscapes (Figure 2 – bottom). With respect to the objective space dimension and to the objective correlation, the autocorrelation lengths are nearly the same.

Our results shed new lights on the definition of a moFiL. According to the hypervolume indicator and to the very elementary neighborhood used in the experiments, the structure of the moFiL is very smooth. The ruggedness of the landscapes depends more on the non-linearity than on the objective space dimension or on the objective correlation. This gives complementary information with respect to [13], which enlighten the importance of objective correlation and objective space dimension on the structure of the Pareto optimal set. Moreover, from the algorithm-design perspective, if we refer on results from single-objective fitness landscapes analysis, a local search based on solution-sets and on the hypervolume should be efficient for the  $\rho MNK$ -landscapes.

### 4.4 Adaptive Walk

In this section, we define an adaptive walk as a first-improvement hill-climbing (HC) algorithm performing on solution-sets. At each algorithm iteration, a random neighboring solution-set is accepted if its hypervolume-value is strictly better than the one of the current solution-set. The walk stops once a local optimum solution-set is found, according to the set-level neighborhood relation. The length of the adaptive walks is studied with a solution-set size  $\mu = 20$ . It reduces the size of the neighborhood structure and then, of the time complexity of the HC algorithm. Usually, it is expected that, when the problem difficulty increases, so is the number of local optima. As a consequence, the length to reach a local optimum becomes smaller.

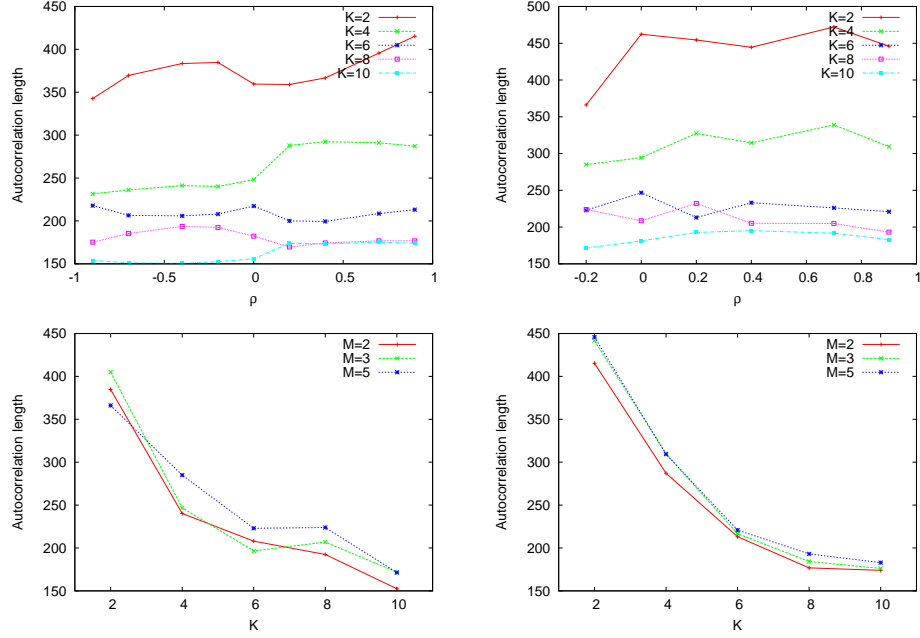


Figure 2: Average value of the autocorrelation length according to parameter  $\rho$  (top left  $M = 2$ , right  $M = 5$ ), and to parameter  $K$  (bottom left  $\rho = -0.2$ , right  $\rho = 0.9$ ). The solution-set size is  $\mu = 100$ .

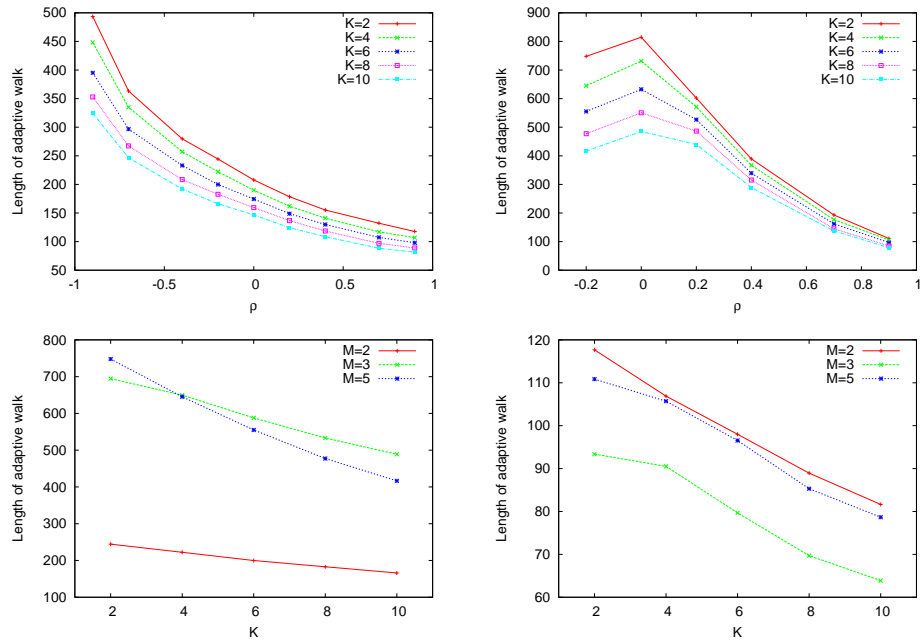
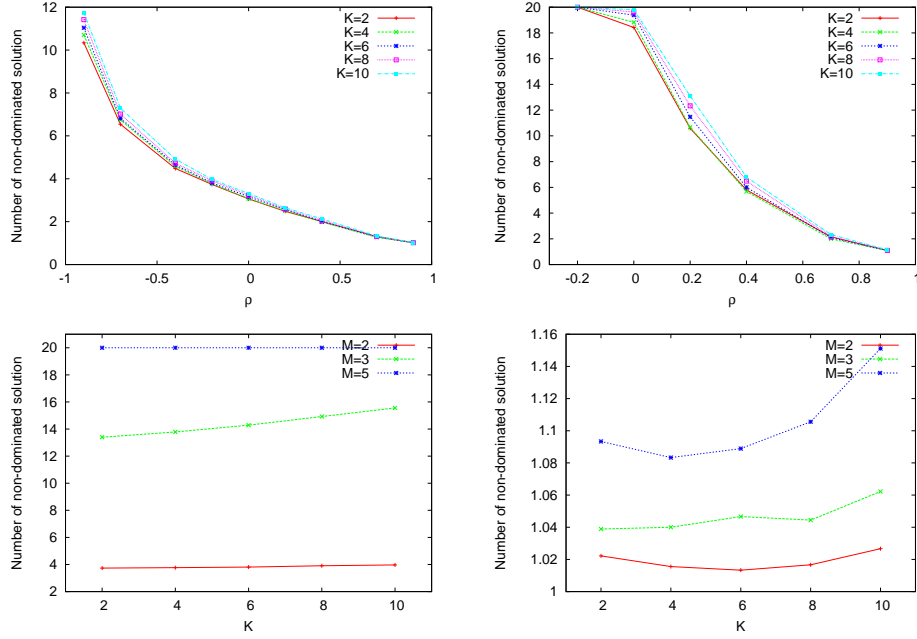


Figure 3: Average length of the adaptive walks according to parameter  $\rho$  (top left  $M = 2$ , right  $M = 5$ ), and to parameter  $K$  (bottom left  $\rho = -0.2$ , right  $\rho = 0.9$ ). The solution-set size is  $\mu = 20$ .



**Figure 4:** Average number non-dominated solutions in the solution-set local optima according to parameter  $\rho$  (top left  $M = 2$ , right  $M = 5$ ), and to parameter  $K$  (bottom left  $\rho = -0.2$ , right  $\rho = 0.9$ ). The solution-set size is  $\mu = 20$ .

Figure 3 shows the length of the adaptive walks according to the  $\rho MNK$ -landscapes parameters. First, as expected, for a fixed objective space dimension and objective correlation, the length of adaptive walks decrease with the non-linearity degree  $K$ . The length is correlated to the difficulty of the problem under study. However, surprisingly, the length decreases when the objective correlation increases, whereas, intuitively, the search becomes easier when the objective correlation increases. A notable exception stands for  $M = 5$ , with  $\rho \in \{-0.2, 0.0\}$ .

In order to explain this result, we need to deeply analyze the set-based HC. Let us note that only non-dominated solutions from the set contribute to the hypervolume. As a consequence, when the number of non-dominated solutions is small, the number of neighboring solution-sets with a strictly higher hypervolume-value is small. In such a case, the length of the adaptive walk should be smaller. This should explain our results. Indeed, according to [13], the size of the Pareto optimal set increases when the objective space dimension and the objective correlation decrease. The non-linearity  $K$  has a low influence on this size. Figure 4 shows the number of mutually non-dominated solutions in the output of the algorithm (*i.e.* in the solution-set local optima). As the size of the Pareto optimal set, the number of non-dominated solutions in the set decreases with the objective correlation, and is nearly constant with the parameter  $K$ . For  $M = 2$ , the maximum size  $\mu = 20$  is never reached. For  $M = 3$ , the maximum size is nearly reached for all correlation values between  $\rho = -0.2$  and  $\rho = 0.0$ .

When there is an equivalent number of mutually non-dominated solutions in the solution-sets of the different landscapes, the length of adaptive walks corroborates the expected property: the larger the size, the ‘easier’ the problem.

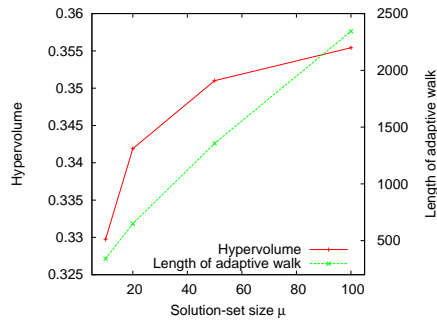
In such a case, like in single-objective fitness landscapes, the length of adaptive walks could be used to estimate the diameter of the basins of attraction of local optima. However, three possible ways could overcome the drawback related to the number of non-dominated solutions. First, it is possible to change the search space or the neighborhood relation in order to consider mutually non-dominated solutions in the sets only. The indicator-based fitness function could also be modified in order to take dominated solutions into account. Second, we can change the definition of the HC algorithm in order to consider the ties in hypervolume-values. At last, when there is a large number of neighboring solution-sets sharing the same hypervolume-value, we can see the fitness landscapes as covered by many plateaus. Then, it could become useful to study the structure of the plateaus more than the solution-set local optima. The decision between these choices has to be made depending on the issues to analyze, and according to the problem and the algorithm under consideration.

At last, as shown in Figure 5, the size of the solution-sets impacts the length of adaptive walks. With the cost of additional evaluations, the quality of the set-based local optimum increases with the solution-set size. Indeed, the length of adaptive walks and the hypervolume-value both increase with the solution-set size, but at different rates. This suggests that a trade-off between cost and quality exists with respect to the solution-set size.

## 5. CONCLUSIONS

In this paper, we formulated a definition of set-based multiobjective fitness landscapes. It is based on a set of solution-sets as a search space, an indicator quantifying the quality of solution-sets as a fitness function, and a set-based neigh-





**Figure 5:** Average hypervolume of the solution-set local optima, and average length of the adaptive walks according to the solution-set size  $\mu$ . The number of objectives is  $M = 3$ , the objective correlation is  $\rho = -0.2$ , and the non-linearity degree is  $K = 4$ .

neighborhood relation. We performed a set-based multiobjective fitness landscape analysis on the multiobjective  $NK$ -landscapes with objective correlation. Our preliminary experimental study shows that tools from single-objective fitness landscapes can directly be extended for analyzing set-based multiobjective search approaches. The relevant features of multimodality and ruggedness has been highlighted for this particular class of problems.

Two difficulties have been pointed out in this work. First, the size of the set-based neighborhood can become very large in comparison with solution-based neighborhood structures. Second, some solutions contained in a feasible solution-set may become dominated by others, so that they do not contribute to indicator-based fitness values for most existing quality indicators. As a consequence, future methodologies will be devoted to an efficient way of sampling the neighborhood, while taking dominated solutions into account. As a next step, we will formalize existing multiobjective search algorithms in terms of set problems based on a set-based neighborhood. Such advances will allow us to analyze the link between the performance and the dynamics of given search methods, together with the main features of multiobjective fitness landscapes. Moreover, we plan to experiment more advance concepts, related to the evolvability, the neutrality, or local optima networks in order to enlarge the understanding of multiobjective problem structures.

## 6. REFERENCES

- [1] H. E. Aguirre and K. Tanaka. Working principles, behavior, and performance of MOEAs on MNK-landscapes. *European Journal of Operational Research*, 181(3):1670–1690, 2007.
- [2] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. TIK Report 286, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, 2008.
- [3] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.
- [4] D. Garrett and D. Dasgupta. Multiobjective landscape analysis and the generalized assignment problem. In *Learning and Intelligent Optimization (LION 2)*, volume 5313 of *Lecture Notes in Computer Science*, pages 110–124. Springer, Trento, Italy, 2007.
- [5] D. Garrett and D. Dasgupta. Plateau connection structure and multiobjective metaheuristic performance. In *Congress on Evolutionary Computation (CEC 2009)*, pages 1281–1288, 2009.
- [6] T.-C. Jones. *Evolutionary Algorithms, Fitness Landscapes and Search*. PhD thesis, University of New Mexico, 1995.
- [7] S. A. Kauffman. *The Origins of Order*. Oxford University Press, New York, USA, 1993.
- [8] J. Knowles and D. Corne. Towards landscape analyses to inform the design of a hybrid local search for the multiobjective quadratic assignment problem. In *Soft Computing Systems: Design, Management and Applications*, volume 2002, pages 271–279, 2002.
- [9] P. Merz. Advanced fitness landscape analysis and the performance of memetic algorithms. *Evolutionary Computation*, 12(3):303–325, 2004.
- [10] J. Mote, I. Murthy, and D. L. Olson. A parametric approach to solving bicriterion shortest path problems. *European Journal of Operational Research*, 53(1):81–92, 1991.
- [11] L. Paquete and T. Stützle. A study of stochastic local search algorithms for the biobjective QAP with correlated flow matrices. *European Journal of Operational Research*, 169(3):943–959, 2006.
- [12] P. F. Stadler. Fitness landscapes. In *Biological Evolution and Statistical Physics*, volume 585 of *Lecture Notes in Physics*, pages 187–207, Heidelberg, 2002. Springer.
- [13] S. Verel, A. Liefooghe, L. Jourdan, and C. Dhaenens. Analyzing the effect of objective correlation on the efficient set of MNK-landscapes. In *Learning and Intelligent Optimization (LION 5)*, Lecture Notes in Computer Science. Springer, Rome, Italy, 2011. to appear.
- [14] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. In *Biological Cybernetics*, pages 63:325–336, 1990.
- [15] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In *International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *Lecture Notes in Computer Science*, pages 862–876, Matsushima, Japan, 2007. Springer.
- [16] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *Lecture Notes in Computer Science*, pages 832–842, Birmingham, UK, 2004. Springer.
- [17] E. Zitzler, L. Thiele, and J. Bader. On set-based multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 14(1):58–79, 2010.
- [18] E. Zitzler, L. Thiele, M. Laumanns, C. M. Foneseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.



